Cross-Layer Optimization: Network Layer Involvement

Oliver Holland, Toktam Mahmoodi, Vasilis Friderikos, and A. Hamid Aghvami

Centre for Telecommunications Research
King's College, University of London
{oliver.holland, toktam.mahmoodi, vasilis.friderikos, hamid.aghvami}@kcl.ac.uk

Abstract—**Cross-layer optimization offers the potential to achieve vast performance gains in wireless communications environments, and even more so in the context of one-to-many services because of the interdependencies among performances at receivers owing to congestion control and reliability requirements. This paper investigates some opportunities for cross-layer optimization involving the network layer. The first such solution, the dynamic selection of the one-to-many data transfer method based on cross-layer information, is shown to achieve considerable bandwidth gains for one-to-many services. The second solution, the cross-layer optimization of routing based on QoS demands, is characterized mathematically. Not only are cross-layer optimizations involving the network layer relevant however, but also are network layer aspects of the facilitation of cross-layer optimization *per se*. Within this context, a means for the transfer of cross-layer parameters using IP extension headers is described.**

Index Terms—**Cross-layer Optimization, Network Layer**

I. INTRODUCTION

CROSS-LAYER optimization represents an important way to improve the performances of future systems, which has until recently been overlooked because of the complexities involved and the feeling in some circles that such approaches deconstruct the original ordered intentions of the ISO/OSI layered protocol stack [1]. However, because the optimal choices for specific layers are often dependent on choices or characteristics at other layers, the consideration of cross-layer parameters cannot be avoided in order to develop systems with anywhere near optimal performance. One way of looking at it is that various fixed cross-layer optimizations are already considered at the design stage of many newly defined communications technologies; it is the extension of such optimizations to greater 'dynamicity' (dynamic in-the-field adaptation), as well as advancing their autonomy and breadth, that are primary objectives of current cross-layer research.

The network layer might be involved in cross-layer optimization in various ways. Not only might it be one of the layers implicated in the cross-layer interaction itself, but it might also offer important functionalities that underpin mechanisms supporting cross-layer optimization *per se*. We attempt to investigate both aspects of network layer involvement in this paper.

This paper is structured as follows. In the next section, a means for the transfer of cross-layer parameters between elements, using advanced network layer capabilities, is described. In Section III, some cross-layer optimizations involving the network layer are introduced. In Section IV, these optimizations are analyzed and simulated in various ways to study their efficiency and characterization. This paper concludes in Section V.

II. MECHANISM FOR TRANSFER OF CROSS-LAYER PARAMETERS USING NETWORK LAYER HEADERS

As well as offering the potential to provide cross-layer optimization solutions, the network layer might also yield facilitators for cross-layer optimization *per se*. To this end, seeing that an over-IP connection is assumed to almost universally apply, a mechanism is suggested here whereby cross-layer information can be transferred between network entities through using IP options [2][3]. This mechanism is fully compatible with all possible other layers, while requiring no mandatory impact on existing protocols.

In IPv6, a Destination Options extension header can be employed to transfer cross-layer information, in which case a new Option (in the Type-Length-Value format) within the Destination Options header would be easily defined for this purpose according to IETF procedures (see Figure 1). IPv6 would be updated where suitable in appropriate hosts (the source and receivers), and the updated implementation would process the new option in the IP layer to provide a specified structure for access to the cross-layer information by layers/processes in appropriate nodes which need it, and are authorized to obtain it. Note that intermediate nodes such as the BS might want to validly append/amend/insert cross-layer information—in such cases these nodes as well as the final destination would be specified in path order in the Routing Options extension header, and a valid alteration of the initial destination address to specify the first such intermediate node requiring cross-layer interaction would apply. Security might conceivably be a concern using Destination Options for such purposes, as any dishonest router could potentially access and alter the information. The Value part of the Type-Length-Value triplet in the newly defined option might therefore be encrypted with a key that is commonly understood only by those nodes that are allowed to input, append/amend, or access cross-layer information. The "Option Type" bits in the extension header would likely be set to 00, thereby still

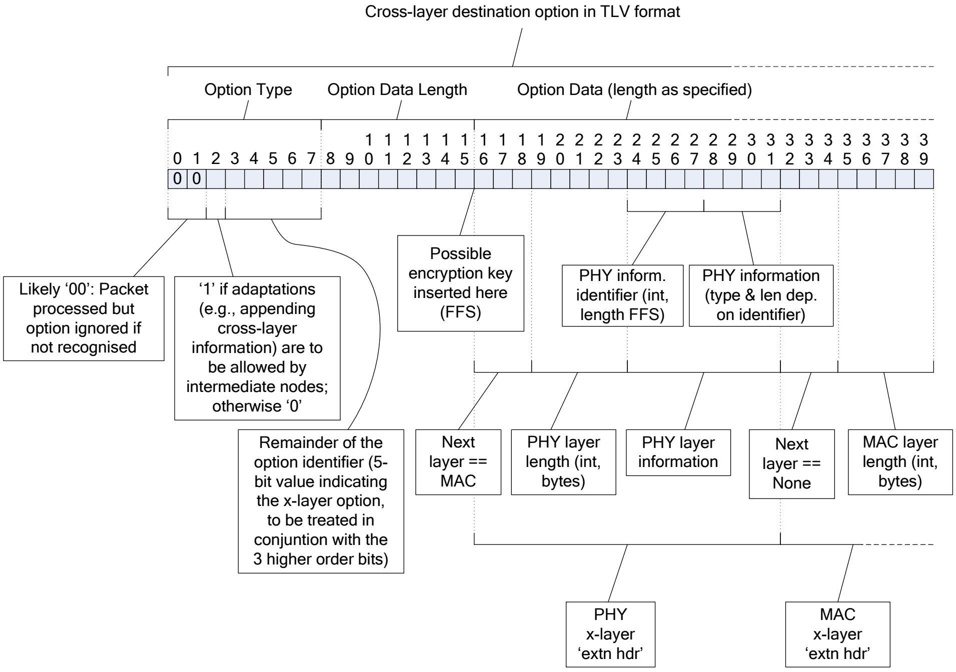


Fig. 1. Suggested format for a novel cross-layer information IPv6 destination option

processing the packet but ignoring the cross-layer information if the IPv6 implementation on the host didn't support the newly introduced cross-layer option. The "Option Change" flag would be set to 1 in cases where it were necessary to allow changes by appropriate intermediate nodes. Remaining questions comprise the exact cross-layer parameters that should be provided, and the utilized information model (i.e., how the information is structured, when it is updated, etc.).

The suggestion in this paper is for structuring of the cross-layer information Option Data to be as in Figure 1. The bits in the Option Data can be viewed as an information hierarchy at three levels. The highest level is the OSI layer that the information corresponds to. Mirroring extension headers, the three highest order bits of the next level are an integer indicating the "next layer" that has information corresponding to it in the header. Note that 3 bits is sufficient for this, giving 8 possible values, one of which (e.g., '000') could indicate that there is no cross-layer information for any other layers in the header. Following this, the next 5 highest order bits indicate the length of the information for the corresponding layer in bytes. Note that 5 bits gives 32 bytes maximum length: it is perhaps for further study whether this is enough, but the structure presented indicates 5 bits as a logical choice to fill the remainder of the byte. At the third tier, the actual information for the layer is grouped into duplets each comprising an integer identifier for the information type (the required length of this field is for further study), which implicitly also indicates the type of the information (e.g., int, float) and the length of the information, followed by the information itself.

IPv6 actually specifies up to two "Destination Options" extension headers, the first of which is processed by the nodes denoted in the "Routing Options" extension header, and the second of which is processed by the final destination node. Hence the first Destination Options header could be used to convey cross-layer information for example to an access point, where the "Routing Options" extension header might specify that access point or an associated node (e.g., the gateway).

Note that a similar method of cross-layer information transfer via IPv4 options might be used. However, this would imply options being processed by all nodes, which may be a considerable burden on routers.

III. SOME CROSS-LAYER OPTIMIZATIONS INVOLVING THE NETWORK LAYER

Here we discuss the optimization ideas that are investigated in this paper.

A. Adaptation of the One-to-Many Data Transfer Method Based on Lower Layer Characteristics

The first cross-layer optimization solution we consider is the use of cross-layer parameters in the (dynamic) selection of the network layer one-to-many transfer method for a one-to-many service, where the transport layer is also somewhat implicated in this choice. A number of benefits result from such an approach: For example, if in an end-to-end multicast the wireless link capacity differs greatly among receivers (with low cross-correlation), the transmission rate of the multicast will be limited to the lowest achievable capacity among the set of receivers. Hence capacity which could otherwise improve the performance of the one-to-many stream or download would remain unused (wasted) in many cells. In such cases, based on cross-layer information, a switch to end-to-end many-unicast might be desired to improve wireless efficiency; alternatively, poorly performing receivers might be siphoned off to unicast connections, thus greatly improving the performance of the remaining multicast. Note that similar important optimization opportunities apply as regards BERs, RTTs and other characteristics varying greatly among the receiver set.

B. Routing in a Wireless Multi-hop/Mesh Scenario

Another considered solution is the cross-layer optimization of routing. In a wireless multi-hop/mesh context, based on traffic QoS demands and concerns for maximizing wireless resource usage efficiency across the system, routes for traffic can be set/updated in a cross-layer context given the characteristics of lower layers. As a simple example of this, if shadowing suddenly becomes severe for one hop of the end-to-end route, and there is an alternative hop available, that alternative hop should be taken to reduce the transmitted power in the area and thus improve spectral efficiency.

IV. ANALYSIS AND SIMULATION

This section covers simulation and analysis of proposed cross-layer optimization solutions.

A. Dynamic Selection of the One-to-many Transfer Method for the Receiver Group as a Whole

Here we first analyze some aspects of the dynamic selection of the one-to-many transfer method for the receiver group as a whole. As is common practice, it is assumed that users receiving the one-to-many service are distributed according to a Poisson distribution with a mean value (per m^2) of λ_d; likewise, it is assumed that the density of users receiving other flows is λ_o. The probability of there being exactly k users per cell is therefore given by [4][5]

$$p(k) = e^{-\lambda 3\sqrt{3}R^2/2}(\lambda 3\sqrt{3}R^2/2)^k / k!\,, \tag{1}$$

where R is the cell "radius" (i.e. the "side" of the hexagonal cell) in m, and λ is either λ_d or λ_o respectively to apply to the one-to-many service or to other traffic users.

As a simple approximation, it is assumed that the spectrum required per capacity in a system is equal among all like channels, where it must be noted that the complications due to elevated transmission power requirements of a common channel are ignored for generality. Hence C_{tot} is taken as the capacity of the cell, and there are n other traffic flow users in the cell which each require a capacity C_o. This leaves C_{tot}-nC_o as the available capacity for the one-to-many service. It is presumed that, if resources allow, it is always preferable to use a dedicated channel twinned with many-unicast to each user for the one-to-many service. This is for reasons such as to ease end-to-end congestion control and the provision of multi-rate capabilities to users, as well as easing power control constraints. Hence the one-to-many service will remain using dedicated channels in the cell until such a situation occurs as it achieves lower than the threshold throughput level C_d, at which point it will switch to a common channel.

Given this approach, the available capacity per one-to-many service receiver is ($Ctot$-nCo)/$k_{download}$ which must be more than or equal to C_d for dedicated channels to be used. Rearranging,

$$k_{download} \le \lfloor (C_{tot} - nC_o) / C_d \rfloor, \tag{2}$$

and the probability of this being met is,

$$\sum_{k=0}^{\lfloor (C_{tot}-nC_o)/C_d \rfloor} e^{-\lambda_d 3\sqrt{3}R^2/2}(\lambda_d 3\sqrt{3}R^2/2)^k / k!\,. \tag{3}$$

Parameter n itself can be described in terms of the probability density $p(k)_{other}$. Hence the probability (3) can be completely expressed as

$$P_{dedicated} = \sum_{n=0}^{\infty}\left[e^{-\lambda_o 3\sqrt{3}R^2/2}(\lambda_o 3\sqrt{3}R^2/2)^n / n! \sum_{k=0}^{\lfloor (C_{tot}-nC_o)/C_d \rfloor} e^{-\lambda_d 3\sqrt{3}R^2/2}(\lambda_d 3\sqrt{3}R^2/2)^k / k! \right]$$

(4)

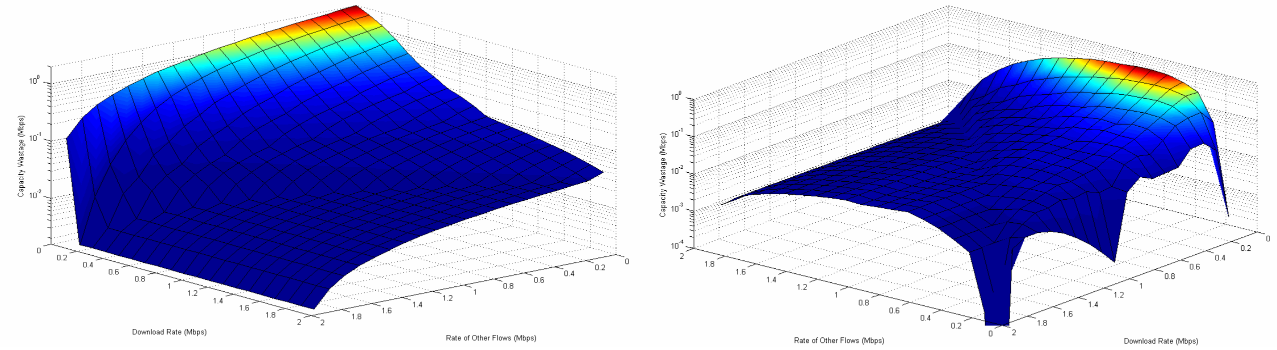
The wasted (Unused) capacity through selection of single-rate end-to-end multicast with common channels throughout, for a large-scale end-to-end one-to-many service, is given by,

$$C_{wasted} = \sum_{n=0}^{\infty}(e^{-\lambda_o 3\sqrt{3}R^2/2}(\lambda_o 3\sqrt{3}R^2/2)^n / n!)(C_{tot} - nC_o - C_d)\,,$$

(5)

where for each iteration in the summation, $(C_{tot} - nC_o - C_d) > 0$ or else that iteration doesn't count.

Figure 2(a) plots this capacity wastage, where C_{tot}=2Mbps, the cell 'radius' R is 200m, λ_o and λ_d are both $4 \cdot 10^{-5}$ users per m^2, and C_o and C_d are both varied between 0.1 and 1.9Mbps in steps of 0.1Mbps. As



(a) (b)

Fig. 2. (a) Average wasted capacity per cell resulting from the selection of multicast using common channels throughout, (b) resulting from the described approach to dynamic channel selection. Both are plotted as a function of C_d and C_o

expected, if C_o and C_d tend to zero, the wasted (unused) capacity in the cell nears C_{tot}.

Next analyzed is the performance of adaptation of the one-to-many data transfer method on a cell-by-cell basis dependent on the concentration of receivers in each cell. The emphasis is on the effect of choosing multicast using a common channel of capacity C_d in the cell, if the required capacity C_d were to not be satisfied using dedicated channels.

The average wasted capacity in each cell of this dynamic selection approach is the wasted capacity $C_{tot} - nC_o - C_d$ given each value of n while satisfying the common channel requirement:

$$C_{wasted} = \sum_{n=0}^{\infty} \left[\begin{array}{c} (C_{tot} - nC_o - C_d) e^{-\lambda_o 3\sqrt{3}R^2/2} (\lambda_o 3\sqrt{3}R^2/2)^n / n! \cdot \\ \sum_{k=\lceil (C_{tot}-nC_o)/C_d \rceil}^{\infty} e^{-\lambda_d 3\sqrt{3}R^2/2} (\lambda_d 3\sqrt{3}R^2/2)^k / k! \end{array} \right], \quad (6)$$

where for each iteration in the outer summation, the condition $(C_{tot} - nC_o - C_d) > 0$ must be met for that iteration to count towards the summation, and of course in the inner summation, each k must be a nonnegative integer.

Given the same parameter choices as in Figure 2(a), this wasted (unused) capacity is plotted in Figure 2(b). As expected, this approach leads to a much better efficiency across the network, especially in cases where the rates of the service and other traffic flows are low.

B. Dynamic Selection of the One-to-many Transfer Method on a Per-Receiver Basis

Here we look at the more intricate approach of dynamically selecting, on a per-receiver basis, which receivers receive the one-to-many service via a separate unicast and which join an existing multicast for the service. For this purpose, a system level simulation platform has been written by us in C++, capable of assessing downlink throughput performances, and configured for a LTE environment.

Under our simulation platform, achieved SINRs were mapped to throughputs using the maximum achievable throughput at each given SINR among the possible modulation and coding schemes for LTE, plotted in Figure 8.1.2.2.1-1 of reference [6]. The assumption was QRM-MLD using ASESS (the blue plot within that Figure). Reflecting common scenarios in [6], the BS-to-BS spacing was chosen to be 500m, and the transmission centre frequency was 2GHz. Mirroring the work reported in Figure 8.1.2.2.1-1 of [6], the transmission bandwidth was chosen as 20MHz, using the suggested OFDMA air interface for LTE. The BS transmission power was set at 50dBm, the thermal noise power was that for a 20MHz bandwidth (-101dBm), and the shadowing standard deviation (std) was chosen to be 6dB, 4dB, or 2dB, resampled in each time step of the simulation. The effective BS height was 30m, reflecting a reasonably sized building, and the height above ground of the mobile was 1.5m. The random walk model comprised the sampling of a change in position in each simulation time unit from a Gaussian distribution (independently for the x and y dimensions) of a specified standard deviation and zero mean. The “Cost 231” path loss model (under the medium-sized city parameterization) was used.

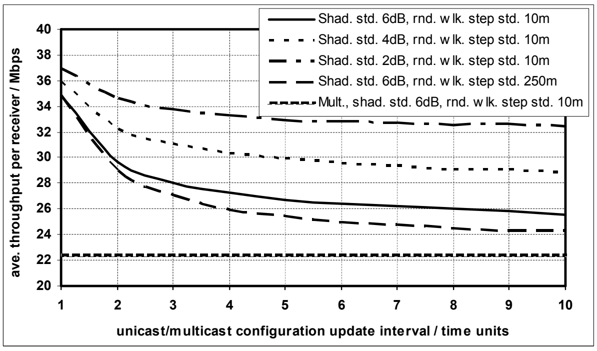


Fig. 3: Average throughput per receiver as a function of the unicast/multicast configuration update interval

Figure 3 plots the average achieved throughput per receiver, considering the effect of load on the source, as a function of the update rate (in simulation step time units) of the unicast/multicast configuration among receivers. Also plotted is the performance of using multicast alone for the service under the stated configuration. Immediately clear is that the average throughput can be improved significantly by updating the unicast/multicast configuration more often, achieving throughput gains of almost ~40% for the 6dB std shadowing and 10m std random walk step plot in Figure 3, and gains of ~45% if the random walk step std is radically increased to 250m. If the situation is more predictable however, represented by shadowing stds of only 4dB or 2dB in Figure 3, the gains of the dynamic unicast/multicast configuration alteration scheme are somewhat less significant. Such gains are still, however, in excess of ~25% for a shadowing std of 4dB, and ~13% for a shadowing std of 2dB. Moreover, the performance improvement of this dynamic selection scheme compared with using multicast alone for the service is extremely significant indeed. Under the assumed parameterization for a 6dB shadowing std, up to ~12Mbps, or ~55% improvement in the average rate of the one-to-many service to receivers, can be achieved.

Figure 4 presents CDFs of throughputs among receivers given a shadowing std of 6dB and a random walk step size std of 10m, for unicast/multicast configuration update intervals of 1 time unit and 10 time units. This Figure further illustrates the potential for capacity utilization improvement through frequently updating the unicast/multicast configuration among receivers, whereby the average performances equate to an improvement for the service from ~1.3b/s/Hz to ~1.8b/s/Hz. Moreover, these plots show that not updating the configuration, or updating less often, has the effect of increasing the uncertainty in the throughput that will be achieved among receivers. This has implications for the types of services that can be reliably supported.

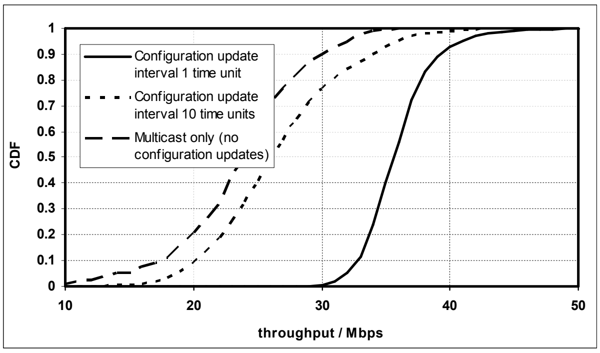


Fig. 4: CDFs of receiver throughputs under a shadowing standard deviation of 6dB and a random walk step-size standard deviation of 10m, for dynamic unicast/multicast selection under configuration update intervals of 1 time unit and 10 time units, and for multicast alone

C. Routing

This section quickly analyses some cross-layer implications for wireless multi-hop/mesh routing. The objective here is to satisfy QoS requirements in an end-to-end sense, hence the concern with the impact that local link characteristics for chosen routes will have on end-to-end performance. First considered are BERs, where the particular interest is with residual BER, i.e. the BER that remains after local link mechanisms have attempted to repair errors. Given a residual BER for each local link *n* of b_n, failure at one or more of the links in the end-to-end path results in the failure of the bit. The probability of failure for the bit is therefore one minus the probability of success occurring concurrently at all links. This probability is

$$P_f = 1 - \prod_{n=1}^{N} (1 - b_n) . \quad (7)$$

This expression might also be considered in terms of PERs/FERs, etc.

Concerning link delays, the delay contributions of all local links must simply be added together to create the end-to-end delay. However, link delays tend to vary considerably hence might commonly be expressed in terms of a delay probability density. The overall end-to-end delay probability density in this case is therefore given as a convolution of all the delay densities for local links [5]. As regards end-to-end delay variability, which might be interpreted as one possible measure of jitter, one simple statistic for this is the coefficient of variation of the resulting end-to-end delay distribution.

Concerning end-to-end capacity, the overall capacity of the end-to-end path is simply the minimum capacity of all local links. Given a CDF of capacity for local link *n*, $F(x)_n$, the CDF of this minimum is simply one minus the product of all CCDFs over links [5]:

$$F(x)_{tot} = 1 - \prod_{n} (1 - F(x)_n) . \quad (8)$$

1) Costs of Chosen Routes

Finally, there is a need to obtain some form of overall value of the "cost" of each possible route, and select a route based on the lowest "cost". One approach is to normalize each factor contributing to the "cost" to 1, then take a weighted average of the contributing factors, where the weighting is in recognition that some factors contributing to the "cost" will be far more important than others (as decided, for example, by network operators). It might also be considered that there will be other factors in this cost, such as the literal financial cost to the user.

Normalizing contributory factors to the cost is perhaps most easily done by introducing maximum acceptable bounds for each contributory factor, where it is noted that these bounds might be different based on the traffic class for example. Indicators such as BER/PER, delay, jitter, and financial cost might therefore be normalized through

$$Cost_i = Cost_{actual(i)} / Cost_{threshold(i)} \quad (9)$$

and capacity might be normalized using the inverse of this. The overall cost for a chosen route can then be obtained as

$$Cost_{tot} = \sum_i \alpha_i Cost_i \quad (10)$$

where α_i is the weighing factor ($0 < \alpha_i < 1$) for the importance of cost factor *i*.

V. CONCLUSION

This paper has investigated some solutions and facilitators for cross-layer optimization involving the network layer. A method for the exchange of cross-layer parameters has been introduced, using IP options headers. Among the solutions analyzed and simulated, it has been shown that significant data-rate gains for one-to-many services can be achieved through dynamically updating the unicast/multicast configuration among receivers. Some aspects of the characterization of network layer routing based on cross-layer QoS constraints have also been analyzed.

ACKNOWLEDGMENT

The work reported in this paper has formed part of the Delivery Efficiency Core Research Programme of the Virtual Centre of Excellence in Mobile & Personal Communications, Mobile VCE, www.mobilevce.com. This research has been funded by EPSRC and by the Industrial Companies who are Members of Mobile VCE. Fully detailed technical reports on this research are available to Industrial Members of Mobile VCE.

REFERENCES

[1] V. Kawadia and P. Kumar, "A Cautionary Perspective on Cross-Layer Design," IEEE Wireless Comm. Mag., Vol. 12, No. 1, February 2005

[2] M. Goncalves and K. Niles, *IPv6 Networks*, McGraw-Hill, June 1998

[3] M Miller, *Implementing IPv6 (Second Edition)*, M&T Books, 2000

[4] C. Walck, "Handbook on Statistical Distributions for Experimentalists," Internal Report SUF–PFY/96–01, University of Stockholm, 2001

[5] A. Papoulis and S. Pillai, Probability, Random Variables and Stochastic Processes, Fourth Edition, McGraw-Hill, December 2001

[6] 3GPP TR 25.814 (V2.0.0), "Specification Group Radio Access Network; Physical Layer Aspects for Evolved UTRA," Release 7, June 2006